CHAPTER

3

MOTHERBOARD ARCHITECTURE, PROCESSORS, MEMORY, AND BIOS

After reading this chapter and completing the exercises, you will be able to:

- Explain various motherboard buses
- Describe how clock frequency affects performance
- Identify common server processors
- Identify various types of memory
- Configure the BIOS and identify common server configuration items

The motherboard provides the system "bus," the transportation medium for data to and from processors, memory, peripherals, and input/output (I/O) devices. Knowing about these server components (and those in Chapter 4) helps you to make educated decisions in the server equipment you procure. The motherboard—sometimes referred to as the system board or backplane—is the "mother of all circuit boards." The processor(s), memory, buses, adapter cards, I/O ports, mass storage—just about every component in the system—are connected directly or indirectly to the motherboard. The various hardware components attached to the motherboard require basic management using a **BIOS (basic input/output system)**, which identifies and confirms correct hardware installation and configuration.

Most readers of this book have had at least some experience with personal computers and might already understand many of the basic hardware components. However, when addressing hardware relating to servers, you must be careful to change perspective. Not only is server hardware more powerful in many respects, but it also includes different features and functions than you might find in a PC.

This chapter outlines the most important factors in assessing server components to help you make informed server purchasing decisions. For budgetary purposes, you might be more interested in building a server on your own instead of purchasing a preassembled server. Although Chapter 6 explains how to install major components, I strongly recommend that you purchase servers as much as possible as pretested turnkey solutions (configured hardware and operating system ready to run right out of the box) from major server vendors. Technicians acquainted with PC workstations might be tempted to use their technical expertise to assemble their own server systems. However, server availability is a high priority, and even the best technician cannot responsibly guarantee the same system uptime as major PC server vendors such as Dell, Compaq, and Hewlett-Packard (HP), which offer server systems with between 99.9% and 99.999% uptime. (Server technicians colloquially refer to this as a certain number of nines. For example, 99.999% uptime is referred to as "five nines" and equals about five minutes of downtime per year.) Some vendors even promise financial compensation if your server fails.

Get on the Bus

The bus is to the server what a highway is to a transportation system. The **bus** provides the data path to and from server components such as the processor and memory on the motherboard, the foundation of the computer. The motherboard attaches to the chassis and includes slots, sockets, and other connections for server components. A foundational architectural factor of the motherboard and its components is the bus width, in bits. Motherboard bus width corresponds to individual data wires that transmit data. The more wires a component such as the motherboard has, the more data it can transmit in a given period of time. Current motherboard data bus architecture is either 32 bits wide or 64 bits wide, which you could equate with a 32- or 64-lane data "highway." Of course, the 64-bit data highway will be able to deliver twice the data in the same amount of time as a 32-bit data highway. As this section will explain, a motherboard includes a front side bus, sometimes a back side bus, and three primary I/O expansion buses: Industry Standard Architecture (ISA), Extended ISA (EISA), and Peripheral Components Interconnect (PCI). The speed of each bus is dependent upon the motherboard clock frequency.

Clock Frequency

Each bus and every device that connects to the motherboard bus depends on the clock frequency of the motherboard. **Clock frequency** (sometimes called the clock speed, cycle, or clock cycle) is the number of times in one second that an electrically charged quartz crystal located on the motherboard vibrates (oscillates). Clock frequency is measured in megahertz, a hertz equaling one cycle per second and "mega" meaning million. If a motherboard has a clock frequency of 66 MHz, then it cycles 66 million times per second. (Most new motherboards have a bus clock speed of 100 or 133 MHz.) The importance of the clock speed is that the processor requires at least one cycle (and usually more cycles) for each instruction that it executes. Therefore, the more times the motherboard clock cycles, the more instructions the processor can perform per second. An 800 MHz Pentium processor has about 800

million opportunities per second to perform an action, subtracting wait states in which the processor uses empty clock cycles to wait for another instruction or hardware function to complete. Other system components, such as buses attached to expansion slots, also depend upon the clock cycle to determine the speed with which they operate.

You can increase the computing speed of the processor by increasing voltage to the processor, thereby improving performance. This procedure is known as **overclocking**, a risky process in terms of overall stability and increased temperature. I do not recommend overclocking production servers, but if you want to experiment (at your own risk) on a spare PC, first check into specific overclocking instructions and precautions at *www.tomshardware.com* and *www.overclockers.com*.

This chapter makes no attempt to explain every type of bus, peripheral, and port, because many simply do not apply to current technology or to servers. (Examples include VESA and MCA buses and game ports.) However, several buses apply to PC servers, in particular PCI buses, which are explained in this chapter.

Chipset Function

In early personal computers, the entire system and its components operated on one bus and ran at 4.77 MHz. This fact seems unimportant until you consider the dramatic changes in bus speeds in the computing industry. Various hardware components require differing bus speeds in order to perform well. Faster components can run without waiting for slower devices to complete their tasks, because the devices operate in an independent bus context. If all components continued to run on a single bus as in early PCs, the resulting bottlenecks would significantly defeat computing efficiency.

Using different buses in the system requires a way to divide the motherboard into separate parts. The **chipset** (see Figure 3-1) is a group of motherboard chips that operate at the same speed as the motherboard clock and provide the boundary that divides one bus from another and controls the flow of bus data. Choosing a motherboard is mostly choosing a chipset.

Figure 3-1 The chipset is identified directly on the chip

Although I do not intend to recommend a specific manufacturer over another, I will direct you to Intel's web site at *http://support.intel.com/support/motherboards/server,* which displays a current listing of server motherboards/chipsets. Intel is the primary manufacturer of PC-based server boards, and regardless of the name on the server case, most PC server vendors use Intel chipsets. Other vendors also make chipsets that are compatible with Intel processors, but again, they tend to be low-cost alternatives and often focus on the desktop computer market. I do not address chipsets geared toward desktop client computers, although a small organization can use a desktop PC as an inexpensive entry-level server. For our purposes throughout the remainder of this book, only systems supporting two or more processors shall be considered servers unless stated otherwise. Also, I do not address other chipsets such as those designed for AMD processors because AMD tends to focus on desktop computers with the exception of the AMD Sledgehammer (see the section on processors in this chapter), which is developing a following in the Linux community. Nevertheless, AMD provides excellent low-cost, high-speed processors and is making significant headway into the desktop PC market. For more about AMD products, visit *www.amd.com* and also refer to Course Technology's *Enhanced A+ Guide to Managing and Maintaining Your PC* (ISBN 0-619-03433-5) by Jean Andrews (*www.course.com/pcrepair*).

Hierarchical Bus

PC-based bus systems use what is known as a **hierarchical bus**, because several buses actually comprise the (collective) "bus," each running at different speeds and with the slower buses hierarchically structured beneath the faster buses. Dividing the bus into the front side bus, PCI bus, and ISA buses allows slower components to operate without negatively impacting the faster components. (There is also a back side bus, which we

will discuss later.) Intel architecture utilizes a North Bridge, South Bridge, and Super I/O chipset (see Figure 3-2) to divide the PCI bus from fastest to slowest and facilitate communication between buses in the order listed:

- **Front side bus**—A 64-bit data pathway that the processor uses to communicate with main memory and the graphics card through the North Bridge chipset. The **North Bridge** chipset divides the processor bus from the PCI bus, and manages data traffic between the South Bridge (see below) and between components on the front side bus and PCI bus. This core bus runs at motherboard clock speed. The front side bus is also known by several other names, including processor bus, memory bus, and system bus.
- **PCI bus**—A 32-bit data pathway for high-speed I/O for expansion adapter cards, USB, and IDE ports. The CMOS (defined later in this chapter) and system clock also connect to the PCI bus. The PCI bus connects to both the North Bridge and the South Bridge. The **South Bridge** separates the PCI bus from the ISA bus. (See more about the PCI bus later in this chapter.)
- **ISA bus**—A 16-bit data pathway for slower expansion adapter cards and the floppy disk, mouse, keyboard, serial and parallel ports, and the BIOS via a Super I/O chip, which mitigates the need for a separate expansion card for each of the aforementioned items. The South Bridge connects to the ISA bus, which is the end of the hierarchical bus chain. (More about ISA follows.)

It might be helpful for you to know the throughput capabilities of the following types of ports and buses, because many times you will need to transfer data through these ports, even as a temporary solution. For example, you might need to transfer diagnostic data from one server to another or to a laptop through a special serial cable known as a **null modem cable**. This type of cable uses special crossed wires to simulate a modem presence, allowing data to travel between the two hosts. In addition, many devices such as UPS systems are now connecting to the server via a USB port instead of a serial port, and new external hard disk storage devices can connect through a FireWire (IEEE 1394) port. Table 3-1 shows the various interfaces and their maximum throughput from the slowest to the fastest.

Table 3-1 Throughput Capabilities of Ports and Buses

Port or Bus	Maximum Throughput
Serial	230 or 460 Kbps with a 16650 UART*
Parallel	500 KBps to 2MBps with ECP**
USB 1.1	12 Mbps
USB 2.0	480 Mbps
IEEE 1394 (FireWire)	200 Mbps (though future specifications will go as high as 1 Gbps)
SCSI-3 (Ultra320)***	320 MBps

* UART is Universal Asynchronous Receiver/Transmitter, a special serial port chip that increases throughput.

** ECP is Enhanced Capabilities Port, a high-speed, bidirectional parallel port.

*** See more about SCSI in Chapter 5.

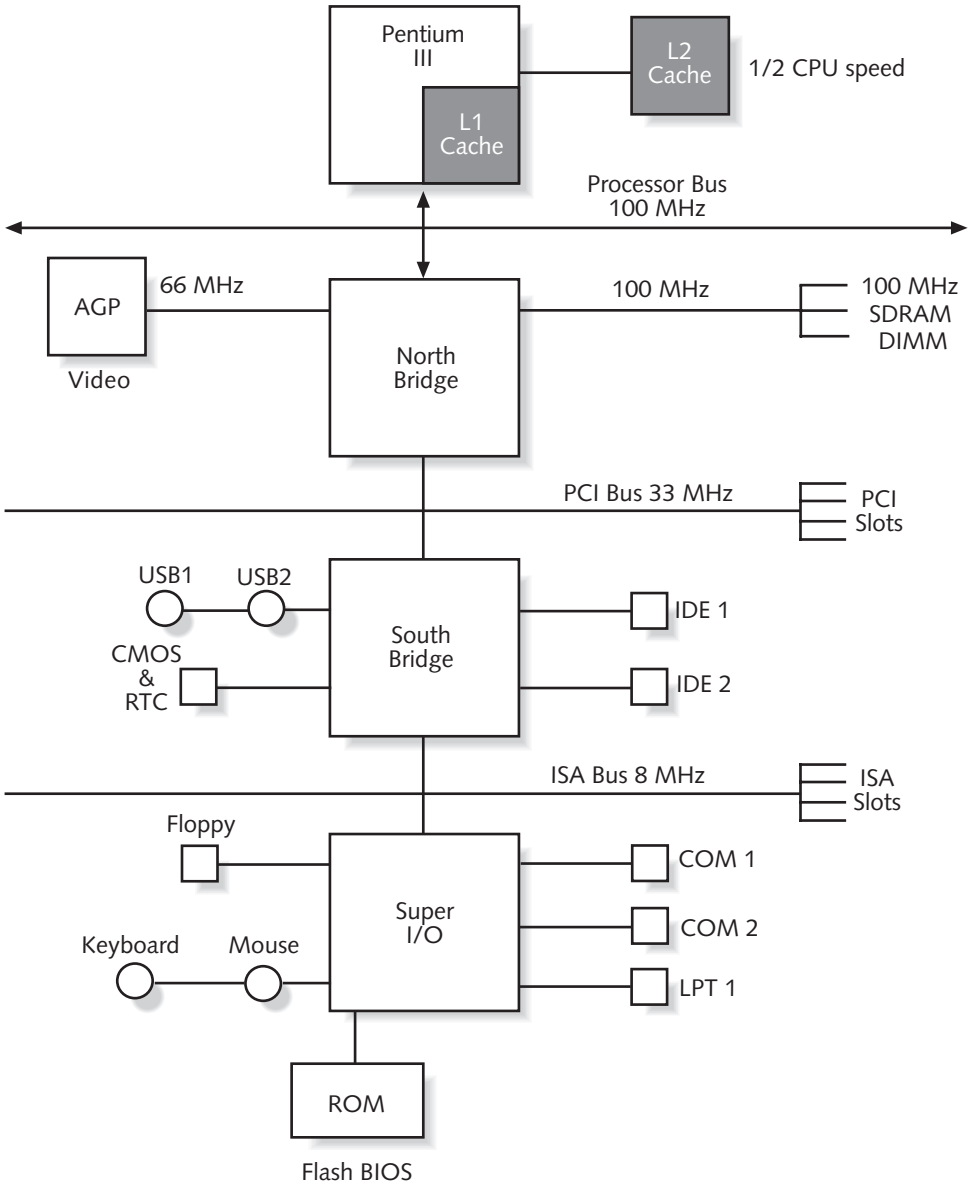

Figure 3-2 Typical North Bridge/South Bridge motherboard architecture

Accelerated Hub Architecture

Intel implements a new hub architecture to replace the tried-and-true North Bridge/South Bridge architecture, which connects the various buses described above through the PCI bus. The **accelerated hub architecture** connects buses to the system bus independently through a dedicated interface to the PCI bus, yielding throughput of up to 266 MBps—twice as much throughput as 33 MHz PCI. The independent buses do not share the PCI bus, thus increasing PCI bus bandwidth available to PCI-connected devices. Also, the hub architecture improves traffic throughput between slower I/O buses and the system bus.

In an accelerated hub architecture (see Figure 3-3), the North Bridge is called the **Graphics Memory Controller Hub (GMCH)**, and the South Bridge is called the **I/O Controller Hub (ICH)**. This architecture allows devices directly connected to the ICH (such as high-speed ATA-66 and ATA-100 disk controllers and USB 2.0 interfaces) much greater throughput.

North Bridge/South Bridge architecture is common on Intel 44*X* series chipsets, such as the 440LX, and the accelerated hub architecture is the current architecture in Intel 8*XX* series chipsets, such as the 840NX.

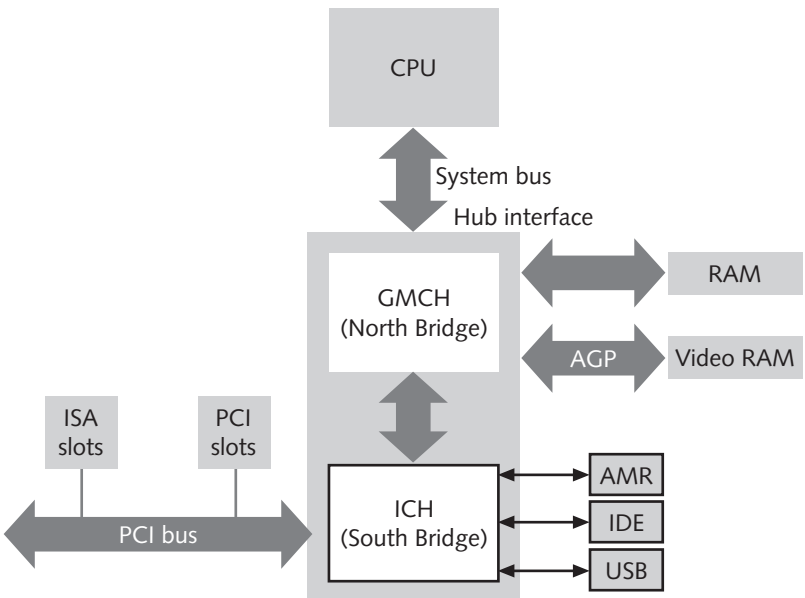

Figure 3-3 The accelerated hub architecture improves I/O traffic between slower I/O buses and the faster system bus

Bus Interfaces

The front side bus, ISA bus, and PCI bus are only useful if there is a way to connect devices to each respective bus. The front side bus uses a slot or socket on the motherboard to connect the processors and memory. The ISA and PCI buses use expansion slots. Although PCI is quickly becoming the only slot available on new motherboards, you might still see the older ISA/EISA slots on some motherboards, so we briefly discuss them as well.

ISA

Industry Standard Architecture (ISA) was originally the AT (advanced technology) bus, developed by IBM (not to be confused with the AT motherboard). Devices connect to an ISA bus through an ISA expansion slot, which is 16 bits wide and accommodates both 16-bit devices and older 8-bit devices. The ISA bus operates at only 8.33 MHz and is capable of transfer speeds up to 8 MBps. This performance seems slow compared to the clock speed of the motherboard. However, most ISA devices (a modem, for example) are even slower, so the slot performance does not hinder performance of the ISA device. You might see some older servers with ISA expansion slots, but newer server motherboards (and workstations or home computers) do not usually include them. Motherboards still include the ISA bus (not the expansion slots) to accept a Super I/O chip as the means to connect slower devices such as the serial and parallel port, floppy controller, keyboard, and mouse.

EISA

The 32-bit **Extended ISA (EISA)** bus provides backward compatibility with older ISA devices and a maximum bandwidth of about 33 MBps. Even though EISA performance is better than ISA, most motherboards do not include EISA buses any more because most manufacturers engineer devices that are compatible with the better-performing PCI bus interface.

PCI

The purpose of the **Peripheral Components Interface (PCI) bus** is to interface high-speed devices with the system bus so that slower devices do not create a bottleneck. For example, an older computer with a 16-bit ISA, 8 MHz video card would create a significant bottleneck for the rest of the system (mostly the processor) because it can only transfer a maximum of 8 MBps, and would use clock cycles that could otherwise be used by other devices. The processor would have to wait for the ISA video card to complete its task before being able to use the system bus. Using the bridge or hub architecture, faster PCI devices can use the PCI bus, reducing or removing the bottleneck. Continuing the video card example, a PCI video card uses a 32-bit 33 MHz card for much faster performance and better throughput at 132 MBps. Other devices such as network cards and hard disk controllers can also take advantage of this improved performance.

As for usability, one of the most significant benefits of PCI is that a Plug and Play operating system such as Windows 95 or later can automatically detect and assign system resources to new devices. Other buses such as ISA required BIOS configuration and/or manual configuration of jumpers and switches on devices. PCI 2.2 is the current specification of PCI, and PCI-X is soon to supercede it (see below).

The most authoritative web site on PCI is *www.pcisig.com*. However, you can only view limited information unless you either pay a fee for each download or pay a hefty annual membership fee. This site is geared toward developers who need to know specific electrical details about PCI so that they can develop products using the PCI architecture.

The PCI standard is compatible not only with PC platforms, but also with Macintosh, Sun, and Alpha platforms using platform-specific chipsets.

PCI-X

PCI-X (**PCI-eXtended**) is actually Addendum 1.0 to the PCI 2.2 specification. The basic advantage to PCI-X is simple: much higher bandwidth and correspondingly higher performance. PCI-X utilizes 64 bits and up to 133 MHz, yielding a maximum bandwidth of 1064 MBps. Devices that are designed according to the PCI-X standard will be able to utilize the full maximum available bandwidth, provided no other processes or devices contend for the same bandwidth. Also, relaxed ordering arranges real-time audio and video instructions in an efficient order instead of the first in/first out (FIFO) method of previous PCI versions.

Relaxed ordering could be a significant advantage if you use servers to deliver multimedia content to the web, for example.

Other efficiency enhancements to the PCI-X bus help to free up bandwidth and reduce wait states, with the net result of a nearly tenfold performance increase over 32 bit, 33 MHz PCI. Motherboard designers divide the PCI-X bandwidth in one of several slot combinations for each PCI-X bus segment: one 133 MHz slot, two 100 MHz slots, or four 66 MHz slots. Table 3-2 shows PCI and PCI-X performance statistics.

Table 3-2 PCI Performance

Data Path Width (bits)	Bus Speed (MHz)	Max Bandwidth (MBps)
32	33	133.33
32	66	266.66
64	33	266.66
64	66	533.33
64	133 (PCI=X)	1066.66 (PCI=X)

Bus Mastering

Most devices utilize the processor to control the flow of information through the bus. As a result, a processor laden with the task of controlling requests from various devices is not as available to process more important productivity functions, slowing down overall performance. In PCI architecture, hardware designers can use **bus mastering** to bypass the processor and directly access memory, resulting in an overall increase in processor performance. Bus mastering is actually a form of direct memory access (DMA) known as first-party DMA. "First party" refers to the device directly controlling memory access, and compares to a third-party DMA transfer using a motherboard DMA controller. Also, bus mastering devices can communicate among themselves over the bus without CPU intervention. Video adapters and disk controllers commonly utilize bus mastering.

PCI Interrupts

Devices issue requests for system resources using an ISA-based interrupt request. An **interrupt request (IRQ)** is an electrical signal that obtains the CPU's attention in order to handle an event immediately, although the processor might queue the request behind other requests. Most devices utilize one of several IRQs on the motherboard. However, there are a limited number of available interrupts, and the number of devices is often greater than the available IRQs. The BIOS utilizes the PCI bus to assign special **PCI interrupts** to PCI devices using the designation INTA#, INTB#, INTC#, and INTD# (sometimes known simply as #1-#4). Single-function PCI cards always receive INTA# according to PCI specifications. Chips or cards with multiple functions can receive assignments for INTB# through INTD# as needed. The PCI interrupts map to one of four corresponding ISA IRQs, usually IRQ 9-12. For example, if you have three single-function PCI cards, they all receive INTA#; however, each device still requires a unique ISA IRQ mapping. Functionally, the result is not that much different than if each device was a standard ISA device in the first place, because each device still receives unique, nonshareable IRQs. The benefit of the PCI interrupt appears when no more ISA IRQ addresses are available. With no more available IRQs, the PCI interrupt utilizes another PCI function known as **PCI steering**, in which the PCI interrupt assigns two or more PCI devices the same ISA IRQ.

PCI Hot Swap

PCI hot swap, otherwise known as **PCI hot plug**, means that you can add, remove, or replace PCI devices without first powering down the server. Note that even though it sounds like you can just take off the cover and rip out a card, many devices and PCI slots require you to follow specific steps. For example, most servers require you to turn off the power to the slot using management software (as with many Dell and HP systems) or using a switch or button (as with Compaq systems) before removing the device (see Figure 3-4). The power switch is often a button located near the actual slot. After you turn off the power to the slot, then remove and replace a card, you can turn the power back on. If the NOS is Plug and Play compatible (such as Windows 2000), it should be able to detect the new device and load (or request) the appropriate drivers. Hot swapping is truly a lifesaver in servers that require 24/7 operation.

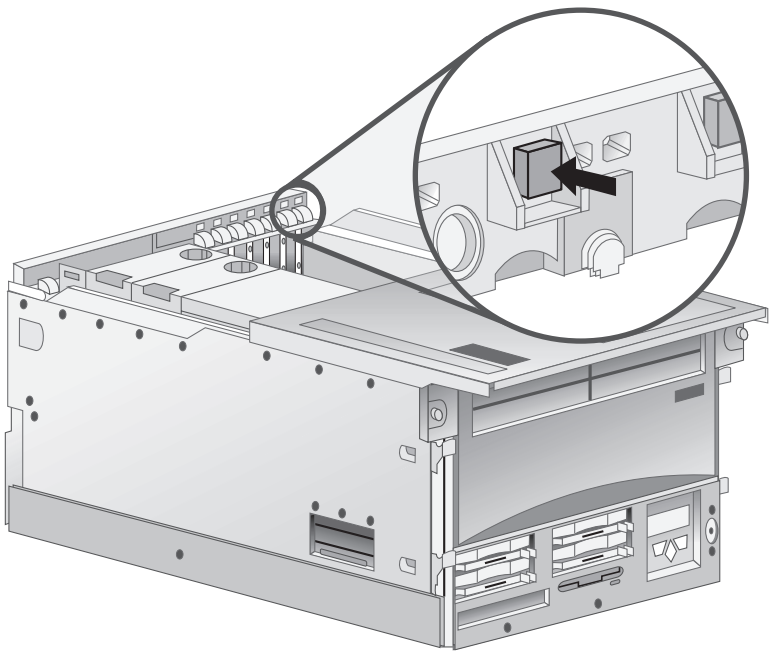

Figure 3-4 To turn off power to a PCI slot, use a button (as shown) or management software

Peer PCI Bus

The **peer PCI bus** is usually a server-specific function that both increases available PCI bandwidth and expands the number of PCI expansion cards from the usual limit of four with a minimal impact on overall system bus bandwidth. This architecture usually involves dual peer PCI buses and two North Bridges, which connect to a primary PCI bus and a secondary PCI bus. PCI expansion slots connect to each respective PCI bus and are either integrated into the

motherboard or installed as add-on daughtercards. Many motherboards use this expanded functionality not only to increase the number of expansion slots, but also to offer flexible PCI bus width and speed. For example, Bus #1 could offer four standard 32-bit 33 MHz PCI slots, while Bus #2 offers two additional 64-bit 66 MHz PCI slots, and devices in both buses can simultaneously access their respective buses. Peer PCI slots allow the administrator to load balance the system. For example, if you have two high-speed network cards for which you expect a great deal of traffic, you could place each one on a separate PCI bus to balance the load. That way, they can each handle I/O without waiting for the other to complete a task on the PCI bus. You can extend the same load-balancing benefits to other devices such as high-throughput SCSI controllers. High-end servers such as the HP Netserver running Windows 2000 Datacenter Server offer up to 32 processors and 96 PCI slots!

Compare peer PCI slots with the bridged PCI bus, in which an additional PCI-to-PCI bridge is inserted below the North Bridge. This only increases the number of available slots, but does not offer better PCI load balancing because all expansion slots actually use a common data path to the system bus. In the peer PCI architecture, separate buses independently communicate with the system bus for more efficient load balancing.

I2O

Intelligent Input/Output (I20) is an initiative to improve I/O performance via an I2O processor and driver model. The I2O driver communicates with the I2O processor, which is located on the device itself, as a separate add-in card, or integrated into the motheboard. Even on the PCI bus, which is designed to relieve traffic from the system bus, frequent PCI interrupts to the processor slow overall performance. With I2O, devices intelligently perform much of the processing function on their own. Also, I2O devices can communicate among themselves when necessary instead of using the processor to manage their communication. The I2O driver utilizes a "split driver" model in which the Operating System Module (OSM) handles I/O interaction between the device and the operating system, and the Hardware Device Module (HDM) manages interaction between hardware controllers and I2O-compatible devices. The I2O specification goes a long way toward developing a common standard that hardware and software vendors can use to simplify and reduce the costly, time-consuming process of driver development. Most I2O-enabled devices are network cards or storage devices because they typically require the highest I/O levels in the system. The I2O specification can work with OS/2, but don't expect to see great strides in I2O technology on the OS/2 platform. Windows NT 4.0 initially offered no participation with I2O; however, Windows 2000 and NetWare 5.x fully support the I2O specification.

The latest I2O specification (2.0) includes several new features, the most significant of which are:

- 64-bit addressing accommodates increased memory capabilities for newer CPUs.
- Hot-plug capability lets you change the adapter without shutting down the entire system.

- Direct memory access (DMA) allows direct access to memory instead of first utilizing the processor. High-speed I2O RAID disk controllers in particular benefit from DMA.

Accelerated Graphics Port (AGP)

The **Accelerated Graphics Port (AGP)** is designed to relieve the system bus and CPU of traffic and processing. Producing graphics is a very complex function, requiring memory usage and significant processing power. In the past, graphics functions used portions of main memory and depended on the main CPU to process much of the graphics load. Later, as video adapters matured, they performed much of their own graphics processing by adding memory chips and onboard processors specially designed for graphics functions. Nevertheless, the graphics card would frequently request attention from the system bus and the processor.

The AGP specification introduced in 1996 utilizes a single AGP slot on standard motherboards, and further relieves the processor and system bus of video burdens. The slot is brown in color and fits only AGP cards—so you can't accidentally insert any other type of card into the slot.

The AGP specification (*www.agpforum.org*) lists a known problem with AGP cards coming loose from the sockets. Because they are set further away from the back of the computer than other slots, there is a tendency for AGP cards to loosen from vibration, as might occur during shipment. When you receive a new system with an installed AGP card, be sure to firmly reseat the card. Some vendors also include an additional AGP retention device.

The initial specification offered both a 1X and 2X mode, representing a performance multiplier of 2, doubling the effective clock speed of a 1X card. Later, a 4X mode appeared; however, most cards are still produced at the 2X speed. There is also an AGP Pro spec, which uses a longer slot and more pins for higher voltage. In November 2000 the AGP 8X was introduced, but at this writing there are no cards available for it. Table 3-3 shows AGP performance statistics.

Table 3-3 AGP Performance

AGP Mode	Effective Clock Speed MHz (AGP Mode x 66 MHz)	Throughput (MBps)
1X	66	266
2X	133	533
4X	266	1066
8X	533	2133

AGP has an immediate and obvious benefit to overall system performance; however, AGP provides the greatest benefit to graphics-intensive computing, such as PC gaming, computer-aided drafting (CAD), graphic design, and other high-end graphics applications. Some servers come with AGP, particularly dual-processor machines that could just as easily serve as high-end graphics workstations. High-end servers normally do not include AGP because it is not necessary and adds a potential point of failure. Manufacturers try to ensure highest availability for servers by not including complex graphics features. You are not likely to be playing PC games on the server, so there is really no need for AGP graphics. High-end servers usually include a motherboard-integrated video adapter at 1024 × 768 screen resolution and only 256 colors. By graphics standards, this is video from the late 1980s. However, its simplicity avoids potential graphics problems on the server. Also, it does not matter from the administrator's perspective that the graphics are unimpressive, because most day-to-day server administration is actually done remotely on a desktop PC workstation (which probably does have snazzy graphics).

PROCESSORS

It is not within the scope of this book to exhaustively describe every processor known to the PC world, starting with the 8088 pioneer of the Intel platform and finishing with the latest Pentium. You should, however, be aware of the characteristics of common processors found in servers today. Another reason for not covering earlier processors is that you are less likely to find older processors in servers, because administrators tend to update server processing power more frequently than desktop workstation processors. This makes sense because business applications and data on servers become larger and more complex much faster than on desktops, and because server performance affects multiple users. Therefore, a server upgrade yields a higher return on investment than a workstation upgrade, which benefits only a single user. Also, this section focuses on Intel processors for good reasons. First, PC servers by definition involve Intel or Intel-based processors. Otherwise, they would be something like a RISC processor. Also, you will find Intel processors in more PC servers. An Intel competitor, AMD, also makes Intel-compatible processors. However, AMD largely aims its efforts at cost-effective desktop PCs and workstations. We will discuss AMD to some degree in relation to servers because its latest product, a 64-bit processor, is aimed at high-end workstations and low-end servers. In fact, Compaq has started to make servers using the AMD processor.

Processor Speed

Processor speed is a measure in MHz of the number of opportunities per second that the processor can execute an action. Recall that each clock cycle represents an opportunity for the processor to do something. The processor architecture design uses a multiplier methodology to provide the processor's speed. For example, a Pentium III 600 MHz processor installed on a motherboard with a system bus speed of 100 MHz uses a multiplier of 6 (100 MHz system bus speed × multiplier factor of 6 = 600).

Cache Memory

Cache memory is a small amount of memory that stores recently or frequently used program code or data, reducing the latency involved in retrieving data from RAM or disk. Cache memory appears in a number of places on the server, including the hard disk, CD-ROM, and processor. Processors use two types of cache memory: L1 (level 1) and L2 (level 2).

L1 Cache

L1 cache is a small amount of memory (usually 32–64 KB) that provides extremely fast access to its data because of its proximity to the processor and because it runs at the same speed as the processor itself—not at the speed of the motherboard. For example, a Pentium III 850 running on a 100 MHz motherboard utilizes an L1 cache that also runs at 850 MHz, not 100 MHz. L1 cache provides an advantage to system performance, because the processor can access data directly from the L1 cache instead of having to fetch the data from memory, which is slower, or from the hard disk, which is painfully slower. Also, if the data in the L1 cache is the result of a processing action such as a complex calculation, retrieval from the L1 cache conserves valuable processor utilization because a recalculation is not necessary. While the size of the cache seems too small to be of any use, it is a great benefit because frequently used chunks of code or data are constantly served from extremely fast L1 cache.

L2 Cache

L2 cache provides the same basic benefits as L1 cache, but it is larger, ranging from 256 KB to 2 MB. In the past, L2 cache was not stored on the processor die, but was instead stored on a separate chip inside the processor housing. This orientation is known as **discrete L2 cache**. The data path used to access the L2 cache was called the **back side bus**, and it ran at half the processor speed. For example, a Pentium III 450 MHz processor utilizes a 512 KB L2 cache running at 225 MHz. Some Pentium III 500's run a 512 KB L2 cache at half the processor speed and some run a 256 KB L2 cache at full processor speed. Most processors after the Pentium III 500 locate the L2 cache directly on the processor die (similar to the L1 cache) and run it at full processor speed. This **Advanced Transfer Cache (ATC)** is 256 bits wide and eliminates the need for a back side bus.

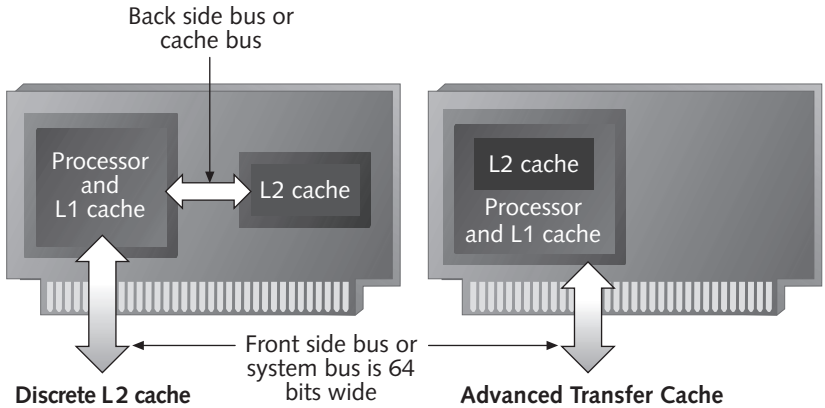

Figure 3-5 Older architectures separate the L2 cache from the processor (discrete L2 cache), while newer processors include L2 cache on the same die (ATC)

Server Processors

In existing servers, you are likely to find processors no slower than Pentium II 233 MHz or Pentium Pro 200 MHz. Servers also utilize an additional processor model not typically found in home or standard desktop PCs—the Pentium II Xeon and Pentium III Xeon (pronounced "zeon"; see Figure 3-6). Xeon processors differ from standard Pentium II or III models in the following respects:

- Type of enclosure
- Cache size
- Cache speed
- Amount of addressable memory
- SMP (symmetric multiprocessing)

Figure 3-6 The Intel Pentium III Xeon

The Xeon uses a Slot 2 single-edge contact (SEC) enclosure and is larger than a Pentium III in order to accommodate the internal board with more L2 ATC memory—up to 2 MB of error checking and correction memory (see more about error correcting code (ECC) later in this chapter). The Xeon uses a 256-bit data path to the L2 cache—a four-fold improvement over the standard Pentium II/III 64-bit data path. On the Xeon, both the L1 and L2 cache can run in parallel, offering simultaneous access and further reducing latency. The Xeon increases the number of fill buffers, the interface between the CPU and main memory, from four to eight, and increases bus queue entries, which hold outstanding bus and memory operations, from four to eight.

The Xeon processor and 2 MB cache add up to 140 million transistors! This fact is more than novelty, because with more transistors comes more heat, so you will have to ensure adequate cooling for this processor.

The amount of memory that the processor can use is a factor of the processor bit width and motherboard chipset. Typically, the processor can address more memory than the motherboard allows. A 32-bit processor can address 4 GB of memory (2^32 = 4,294,967,296 bytes, or 4 GB). In the home desktop and corporate workstation, you are unlikely to find a motherboard with physical space and chipset design to allow for this much memory. Server processor design, however, is changing to allow substantial amounts of addressable memory by modifying the motherboard and/or chipset. Intel Pentium II Xeon and later processors let the processor utilize 36 bits to address memory using Intel's

Physical Address Extension (PAE) feature, allowing up to 64 GB of addressable memory (2^36 = 68,719,476,736, or 64 GB). While even 4 GB sounds like an immense amount of memory (and it is), large, real-time server applications such as online transaction processing (OLTP) and e-commerce require large amounts of data to reside in RAM for fast access.

A Pentium III motherboard configuration accepts either single or dual processors. A Xeon SMP configuration can use up to four processors, though by adding another processor bus (often called a **mezzanine bus**), eight processors are possible, and some manufacturers engineer buses that can use up to 132 processors (though four or eight is more common).

A dual-processor system is known as "2-way," four processors as "4-way," eight processors as "8-way," and so on.

Except for the differences outlined earlier, Xeon processors at the core are no different than Pentium II/III brethren of the same speed. If your server needs do not include significant caching power and more than two processors, you can save a substantial amount of money by using the fastest available Pentium III instead of the Xeon.

Characteristics of Intel Pentium server processors appear in Table 3-4.

Table 3-4 Intel Server Processors

Processor	Processor Speed (MHz)	L1 Cache (KB)	L2 Cache (KB)	System Bus Speed (MHz)
Pentium Pro	200	16	256, 512, 1 MB	60, 66
Pentium II	233, 266, 333, 350, 366, 400, 450	32	256, 512	66, 100
Pentium II Xeon	400, 450	32	512, 1 MB, 2 MB	100
Pentium III	400, 450, 500, 533, 550, 600, 650, 667, 700, 733, 750, 800, 850, 866, 933, 1 GHz, 1.3 GHz	32	256, 512	100, 133
Pentium III Xeon	550, 600, 667, 733, 800, 866, 933, 1 GHz	32	256, 1 MB, 2 MB	100, 133

Notably absent from Table 3-3 and the general discussion to follow is the Celeron processor, because it is designed for the low-cost home PC market. Intel reduces the cost by utilizing a smaller cache and cheaper packaging, although the core Celeron II/III is the same core as the basic Pentium II/III. Also absent are the Classic (original) Pentium and the Pentium MMX because it is unlikely that you will find these in servers—although MMX video technology is still present in server processors. The Pentium 4 processor offers significant performance benefits over the Pentium III; however, it is not capable of SMP and is geared toward the high-end workstation or demanding home user. Intel plans to release an SMP-capable Pentium 4 Xeon (code name "Foster") at 1.7 GHz.

64-Bit Processors

As with motherboard buses and adapter cards, the bit width on a processor correlates to the amount of data that it can transmit. Each bit corresponds to a wire connector through the socket or slot for data transmission between the processor and the motherboard. Most Pentium processors function internally at 64 bits, and then the data results are passed on to the 32-bit external bus interface.

Most server processors use 32-bit bus interfaces, but new processors from Intel and AMD are 64-bit processors both internally and externally. Although these processors are not in final form at this writing, they are likely to affect the future of PC server computing very soon.

Intel Itanium

The 64-bit Intel Itanium using Intel's IA-64 technology represents a departure from the 32-bit x86 Intel architecture, and performs optimally with 64-bit operating systems (Windows 2000 will be ported to 64 bit) and applications. Co-developed with HP, the Itanium depends upon new compiler technology. (A **compiler** translates a high-level programming language into the lowest language the computer can understand, machine language.) In addition, 32-bit applications running on the Itanium processor utilize the Itanium's hardware emulation to adapt the 32-bit instructions for the 64-bit architecture. Because of the translation process, 32-bit applications will usually run more slowly on the IA-64 than on fast 32-bit Pentium III Xeon processors. The Itanium processor runs on Intel's upcoming 460GX chipset.

One of the reasons many large organizations will migrate to the Itanium 64-bit platform (IA-64) is not so much the core processor speed as the 64-bit memory addressability. With 64 bits, the processor can address up to 18 billion GB (2^64 = 18,446,744,073,709,551,616 bytes). This seems like an absurd amount of memory, but at least it doesn't appear as if there will ever be a memory ceiling again, and memory-hungry applications such as databases will make good use of any available memory. In addition to a large L2 cache, the Itanium also supports a 2 or 4 MB Level 3 motherboard cache (much like L2 cache before ATC).

The IA-64 architecture uses Explicitly Parallel Instruction set Computing (EPIC), allowing the processor to simultaneously process as many as 20 operations. New motherboard designs will take advantage of IA-64 architecture to also allow handling of up to 64 bits of data at a time. This type of functionality will be an especially powerful feature when applying 64-bit processing to encryption schemes such as RSA encryption/decryption. Intel estimates that an IA-64 processor will outperform the fastest RISC-based processors by a factor of eight or more.

AMD Sledgehammer

The AMD Sledgehammer (built on the AMD Athlon core) is also a 64-bit processor, but there is otherwise little similarity between the Itanium and the Sledgehammer (or "Hammer"). AMD decided to extend Intel's original x86 architecture in the Hammer design with AMD's new x86-64 architecture. In fact, 32-bit operating systems and applications can run on the Hammer without complicated hardware emulation, resulting in minimal performance overhead. This strategy could prove to be a wise marketing move for AMD, because few enterprises will be able to switch all operating systems and applications to 64-bit overnight. The Hammer allows organizations to gradually merge 64-bit functionality into their existing framework. Unfortunately, Microsoft has committed to creating 64-bit operating systems and applications only for the Intel IA-64 platform, not the AMD x86-64 platform. The Linux community, however, already has a 64-bit version of Linux under way, and Sun Microsystems has also announced that it will port Solaris UNIX to the x86-64 platform. AMD is developing a new Lightning Data Transport (LDT) system bus with throughput as high as 6.4 GBps.

MEMORY

Many types of memory have been available in PCs and servers over the past few decades, but this section describes only the types of memory most likely to be found in servers today. Desktop PCs and servers share many of the same memory characteristics, but servers often have additional memory features, such as registered memory and ECC memory. Several **dynamic RAM (DRAM)** memory chips are installed on a printed circuit board (PCB), which is collectively referred to as a module. DRAM is dynamic random access memory—referred to as dynamic because the information requires continuous electrical refresh, or else the data can become corrupt or lost.

SIMM Modules

You will probably find SIMMs (single inline memory modules) only on older servers. The original SIMM was 8 bits wide (plus one optional parity bit) and used 30 pins to connect to its slot on the motherboard. However, you won't find these older modules in servers. Instead, you are more likely to find SIMMs with a data path of 32 bits (plus four optional parity bits) with 72 pins (see Figure 3-7). The physical SIMM module has gold or tin contacts at the bottom. Although the contact appears both on the front and back, it is really a single contact (hence the *single* in SIMM).

Be careful not to identify a memory module as a SIMM because it has memory chips on one side only or a DIMM (see later in this chapter) because it has memory chips on both sides. It is the contacts on the bottom that differentiate a SIMM from a DIMM.

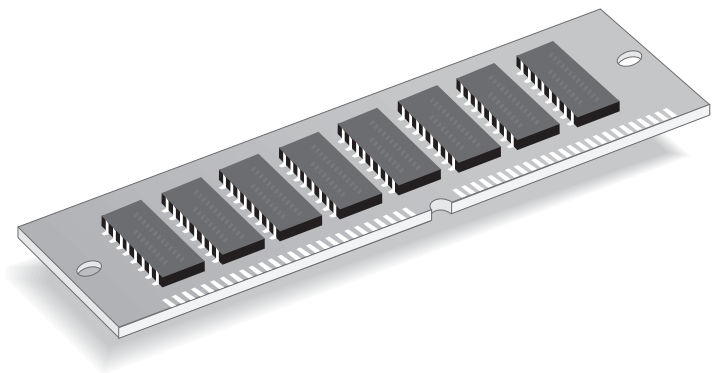

Figure 3-7 A 72-pin SIMM

SIMMs are rated according to the time it takes to retrieve data from memory in nanoseconds (ns), which is one billionth of a second. The typical SIMM is 50 ns, 60 ns, or 70 ns, and older SIMMs can be 80 ns. Even memory running at 60 ns is currently considered slow in light of faster SDRAM DIMMs, described later in this chapter.

EDO

SIMMs and early DIMMs (see next topic) are known for the Extended Data Out (EDO) RAM technology (sometimes called "hyper page mode"), which relates to locations in memory known as memory addresses. A memory address references rows and columns. Instead of providing only the exact location requested, EDO can send the entire row address so that subsequent references to the same row require only a column lookup, saving time. This functionality is the same as an older technology known as Fast Page Mode (FPM) RAM, and also adds the ability to eliminate a 10 ns delay prior to issuing the next requested memory address.

DIMM Modules

DIMMs (dual inline memory modules) dramatically improve memory performance over SIMMs by expanding the module to 64 bits (nonparity) or 72 bits (parity or ECC) using 168 pins (see Figure 3-8). The contacts on both sides of the module are separate (hence the *dual* in DIMM). Recall that the more bits available for the data, the more data that can be processed in a given period of time. Because a DIMM uses 64 bits (instead of 32, like a SIMM), it yields a performance increase.

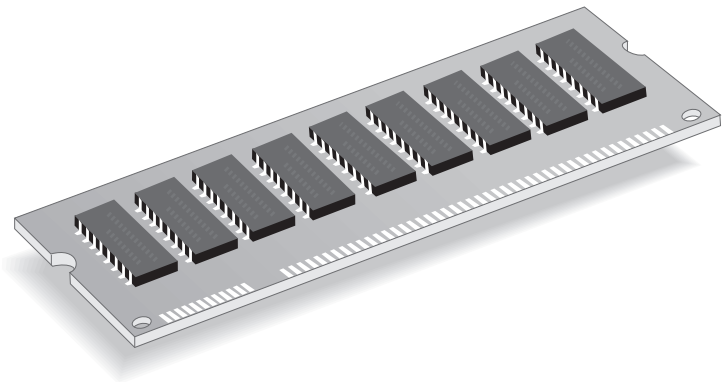

Figure 3-8 A 168-pin DIMM

SDRAM

Closely associated with a DIMM is **SDRAM (synchronous dynamic RAM**), because a DIMM is the physical platform of SDRAM. SDRAM removes the FPM and EDO DRAM signal-controlled bottleneck that emerged as buses faster than 66 MHz appeared. At 60–80 ns, the processor would request information faster than memory could serve it over the bus. SDRAM operates at clock speed; if the system bus is 100 MHz, then SDRAM matches that frequency, which functionally operates at about 10 ns. SDRAM memory is referred to with a PCXXX, where XXX is the bus speed for which the memory is designed. For example, PC100 refers to memory designed with a rated speed for use in 100 MHz motherboards. However, manufacturers actually make SDRAM run at 125 MHz on a 100 MHz bus for added margin because of extremely tight nanosecond timing. SDRAM is also available at 133 MHz at about 7.5 ns. Expect even faster SDRAM to match increasing speeds of newer buses. To identify the speed in nanoseconds upon visual inspection of the memory chip, look at the digits at the end of the product number. You should see 10 for 10 ns, 8 for 8 ns, and so on. Table 3-5 shows common SDRAM speeds.

Table 3-5 SDRAM Speed

Speed in ns	Manufactured Speed in MHz	Rated Speed in MHz
15	66	PC66
10	100	PC66
8	125	PC100
7.5	133	PC133

RDRAM

RDRAM (Rambus DRAM) is an invention of Rambus Technology. Rambus does not actually manufacture memory, but it developed the technology and charges royalties against memory manufacturers. RDRAM memory chips fit on a narrow, 16-bit-wide RIMM memory module. (RIMM is not an acronym; it's a Rambus-patented name). RDRAM provides extremely fast 800 MHz internal clock speed on a 400 MHz bus, because data is transferred on both the leading and trailing edge of each clock cycle. This adds up to 1.6 GB throughput (16 bits × 800 MHz / 8 = 1.6 GB). Intel has expressed the most interest in RDRAM, making it the memory of choice in the 820 chipset for PC desktop platforms, the 850 series for Pentium 4 platforms, and the 840 chipset for high-end workstation and server platforms. However, because of the licensing royalty and tight production tolerances, other chipsets (such as AMD-based chipsets) avoid RDRAM, preferring DDR SDRAM instead (see next topic). The RDRAM data path must travel through each RIMM from beginning to end, which adds a delay when data exits the modules. Compare this to DIMMs, with parallel connections to the motherboard, which allow independent data throughput for each DIMM. Because of the unique data circuit of RDRAM, empty RIMM sockets must be filled with a C-RIMM, a device that has no memory but provides continuity to complete the memory data path. The RIMM is uniquely identifiable because you cannot see the actual memory chips, as with other memory types. Instead, an aluminum sheath known as a "heat spreader" covers the RDRAM to help diffuse high heat levels brought on by the fast access and transfer speeds.

This fact won't help you run your servers, but it's interesting that RDRAM has been around long before its implementation in the PC—it started as proprietary memory for the Nintendo 64!

DDR SDRAM

Double data rate SDRAM (DDR SDRAM) is the next generation of SDRAM, and also uses a 64-bit DIMM with future plans for a 128-bit DIMM. DDR SDRAM (or SDRAM II), like SDRAM, is synchronous with the system clock. However, DDR SDRAM transfers data twice per clock cycle, similar to RDRAM, but at a lower cost because DDR SDRAM is an open standard charging no royalties. If the bus is 133 MHz, DDR SDRAM transfers data at 266 MHz. In addition, it retains the data pathway of DIMMs, offering faster data transfer from the actual DIMM to the bus with parallel construction, as opposed to the continuity requirement of RDRAM.

Table 3-6 summarizes the memory technology of the various types of RAM.

Table 3-6 RAM Memory Technologies

Memory Technology	Calculation of Throughput	Data Throughput
RDRAM	16 bits × 400 MHz /8	800 MBps
RDRAM	16 bits × 800 MHz /8	1.6 GBps
SDRAM on 100 MHz bus	64 × 100 MHz /8	400 MBps
SDRAM on 133 MHz bus	64 × 133 MHz /8	532 MBps
DDR SDRAM on 133 MHz bus	64 bits × 266 MHz	1064 MBps
DDR SDRAM on 166 MHz bus	128 bits × 332 MHz	2656 MBps

Interleaving

Interleaving allows memory access between two or more memory banks and/or boards to occur alternately, minimizing wait states. For interleaving among banks on the same board, you must completely fill the first bank, and then completely fill the second bank with memory that is identical in size and speed. For example, if you have two banks of memory with four slots each, Bank A and Bank B, and Bank A has 256 MB RAM in each slot totaling 1 GB, then Bank B must have exactly the same memory configuration (see Figure 3-9).

If you use larger memory configurations, such as with servers that have separate dedicated memory boards, you can interleave not only among banks on a board, but also among the boards. This configuration also requires you to configure RAM pairs identically. To interleave boards, each pair on one board must exactly match the corresponding pair on the other board (see Figure 3-10). See more about your vendor's specific interleaving requirements. For example, HP has several other considerations for their Netserver Lxr8000 servers (*www.netserver.hp.com*—search for board-to-board interleaving)

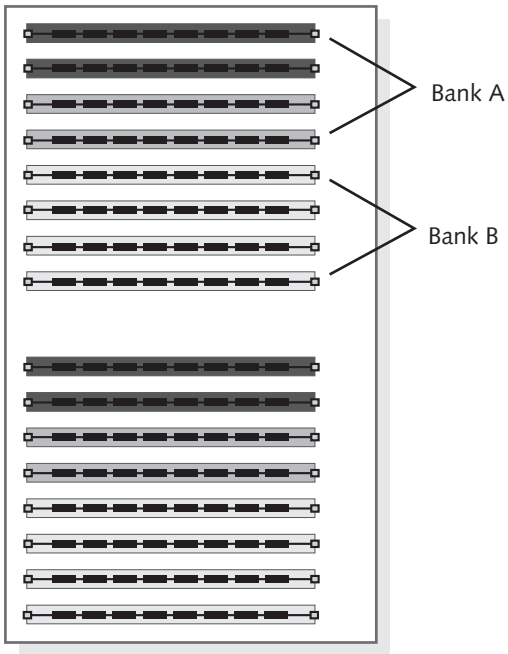

Figure 3-9 Interleaving between banks of memory—Bank A and Bank B must be filled identically

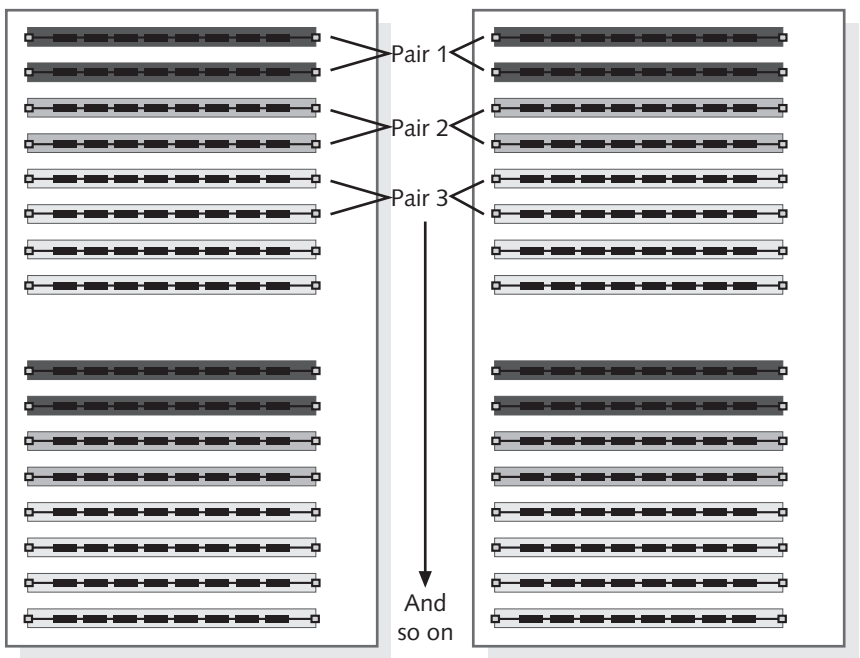

Figure 3-10 Interleaving between boards of memory—corresponding pairs between boards must be filled identically

Interleaving configurations are described in an X-way format, where X is the number of interleaved banks in use. For example, if you have two memory boards with two memory banks, you have four-way interleaving (2 boards × 2 banks = 4-way).

Buffered and Registered Memory

Buffered memory is a function of FPM or EDO memory, and is an older memory technology. The purpose of both buffered and **registered memory** (which is more common in current computers) is to re-drive (amplify) the signal entering the module. Buffered or registered modules, which have a synonymous function, also assist the chipset in handling the larger electrical load when the system has a lot of installed memory, allowing the module to include more memory chips, which is one reason that servers often use registered modules. (With desktop PCs, SIMMs are likely to be unbuffered, because the chipset manages the buffering function.) You find registered memory on servers or high-end workstations, but rarely on a desktop PC. Registered memory also enacts a deliberate pause of one clock cycle in the module to ensure that all communication from the chipset arrives properly. Registered memory is useful on heavily loaded server memory, and was designed for SIMMs containing 32 or more chips.

Error Correcting Code (ECC)

Servers commonly use **error correcting code (ECC)** SDRAM. Although error correction is more expensive and involves a slight performance penalty, it is well worth it on a server, where data integrity is critical and other high-performing system components help make up for memory latency. ECC calculates check bits and appends them to the data during memory writes. For memory reads, ECC decodes the appended check bits and compares the write and read check bits. If there is a discrepancy in the check bits, then an error has occurred, and the NOS can be notified. If only a single bit error occurs, ECC can correct the error, but ECC cannot correct the more rare 2-, 3-, or 4-bit errors (multiple bit errors).

Some complex forms of ECC can detect and correct multiple bit errors, but at this point they are not common.

To find out more about specific memory modules, or to order memory for your servers, check the following web sites:

- *www.crucial.com*
- *www.micron.com*
- *www.kingston.com*
- *www.pny.com*

CMOS, BIOS, AND POST

The computer needs to have a way of finding its bearings—that is, it requires a means to locate, identify, and configure the various hardware components in the system. Hardware settings apply to two basic stages: first when you turn on the system, and second when the NOS loads. The NOS detects and/or applies configured resources to system hardware. This section addresses how the CMOS, BIOS, and POST relate to configuring and detecting hardware when the system is powered on.

CMOS

The **CMOS** is a complimentary metal oxide semiconductor that includes a small amount of memory, the purpose of which is to store the BIOS settings such as the boot order (floppy, CD-ROM, hard disk, and so forth), hard disk configuration, power management settings, and more (see next topic). The CMOS can store data for as long as power is available. The power supply provides power when the system is on, and a small, nonrechargeable, metal oxide battery (similar to a watch battery) supplies power when the server is off. Unlike desktop PCs, servers are usually powered on continuously except for regular maintenance, hardware upgrades, and troubleshooting when you must necessarily power off. Typical batteries can last for more than five years on desktop PCs, and somewhat longer on servers since the battery charge naturally dissipates as a matter of time instead of actual use. Server replacement might precede battery replacement, so sometimes battery life is not an issue. However, you should stock a few of the most common batteries in use on your servers just in case. You identify the specific battery in use by reading the identification stamped into the battery surface. Batteries near the end of their life usually lose time on the real time clock, so you should replace batteries on systems with slowing time.

Older systems copied BIOS settings into a small portion of main memory for faster access. This was known as shadow RAM, which is no longer implemented.

BIOS

The **Basic Input/Output System (BIOS)** is a series of software programs that is the lowest-level interface between the hardware and the operating system. The BIOS programming is stored on a **flash BIOS** memory chip, also known as **EEPROM (electrically erasable programmable read-only memory)**. The administrator can configure the BIOS programming to suit his or her needs and preferences, and the configuration is stored in the CMOS, which is powered by a small battery that retains the settings even when the power fails or is turned off. As its name implies, the BIOS is a series of input and output configuration settings for peripherals, adapters, and on-board components. Phoenix Software, Award Software, and American Megatrends Inc. (AMI) create most base BIOS programming, though individual server manufacturers often add modifications to provide functionality with their specific hardware. Phoenix

Technologies acquired Award Software in late 1998, so a newer system will usually use either a Phoenix or an AMI BIOS. The BIOS controls all of the hardware on the system board and acts as a bridge for various NOS hardware drivers.

We generally consider the BIOS to be only the programming stored in flash BIOS memory, but in reality, it's also collectively on various other hardware and adapters, such as a network or sound card. SCSI controllers, for example, often have their own BIOS programming.

Accessing CMOS Settings

Most of the time, when you turn on the power, the display tells you a specific key or keyboard combination to press in order to access the BIOS settings in CMOS. Typically, this is F1, F2, Esc, Del, or some combination of Ctrl+Alt (such as Ctrl+Alt+Shift, Enter, Esc, or S). Compaq computers usually use F10. A simple instruction such as "Press F2 to enter settings" often appears on the screen. Several manufacturers display a manufacturer-promoting splash screen that would prevent you from seeing such instruction, although pressing the Esc key often removes it.

In the unusual instance that the display does not indicate a method for accessing the BIOS, you can trick the system into allowing you to access the BIOS by pressing and holding virtually any key immediately after you power on the computer. The BIOS will often interpret your action as a keyboard problem and provide an opportunity for you to access the BIOS.

When the system powers on, a procedure known as the **POST (power-on self-test)** verifies functionality of motherboard hardware. If the settings do not match, one or more beeps occur. Check system documentation to interpret the meaning of beep codes, which usually also accompany an on-screen error notification code. During the POST, if a device has its own BIOS such as a video card or SCSI card, the POST allows the device to perform its own diagnostics and then resumes when the diagnostics are finished. The POST checks the following:

- Video card and monitor
- CPU stepping (specific incremental version of the CPU)
- CPU model and speed
- BIOS version
- RAM
- Keyboard (which it enables)
- Various ports such as USB, serial, and parallel
- Floppy and hard disk drives
- Disk controllers using separate BIOS
- CD-ROM or DVD-ROM

- Sound cards
- Operating system (which it finds and loads)

Sometimes you want to see exactly what the POST is displaying on the screen, but it often blinks by very quickly. To freeze the screen, press the Pause key.

Protecting the CMOS

Protecting the CMOS for both the server and workstation is an important security precaution. For example, anyone with physical access to the server could access the BIOS settings in CMOS to ensure that the system can boot from a floppy disk. Then, after booting from an MS-DOS or Windows 98 boot floppy, he or she could gain access to local hard disks and steal, alter, destroy, or otherwise damage data or the operating system. Protecting the CMOS first involves physical security (as addressed in Chapter 2), and then applying a password to the CMOS. The BIOS menu system is usually easily navigable, and you should be able to locate where to designate a password for the BIOS settings in CMOS. The CMOS usually includes two levels of password protection: a password to access and change the BIOS configuration, intended to prevent the curious from viewing the CMOS settings, and a password to boot the system. After setting the password(s), be sure to record and store them in a secure location.

In the event that you cannot find a server's CMOS password, you have no choice but to reset the CMOS, which clears password settings in addition to any configuration settings. Reset the CMOS in one of two ways. First, use jumper pins to short the battery circuit to the CMOS. To find the exact jumper pins, refer to the motherboard manufacturer's manual. In absence of a manual, you can also search for labeling on the circuit board. Usually the label is something like CPW (clear password), RPW (reset password), or the unmistakable "Short here to clear CMOS." Without battery power, BIOS settings drain from the CMOS, and the password resets to the original null (none) password setting. Second, you can simply remove the battery for a few seconds to clear the BIOS configurations, and then replace it. Either way, when you reboot, there will be no password required to boot or access BIOS settings.

Some motherboards have capacitors that retain electricity and could continue to power the CMOS for hours, even without a battery. In this case, you might have to leave the battery out for a day or so. To resume POST, press the spacebar.

Whenever possible, be sure you first back up the BIOS settings either by writing them down or using a utility (such as CMOS.ZIP) that can print settings or save them to a floppy disk. Once you reset the CMOS, the recorded settings can prevent guesswork in reconfiguring the BIOS settings.

If you clear the BIOS settings, the operating system, hardware devices, or system preferences such as power conservation settings might not function as expected. Although a BIOS maker provides the basic BIOS, the server manufacturer has likely shipped the computer with default settings. Without these default settings, items such as the hard disk might not be accessible, making a boot impossible. After resetting the CMOS (or after you replace a dead battery), navigate the menu system to locate a setting that restores default settings to get back to a workable starting point.

Fortunately, the BIOS interacts with several hardware components to automatically configure settings, as in the hard disk configuration and memory detection.

Common BIOS Settings

It would be impractical to list all common BIOS settings from all manufacturers. A general list of configuration categories usually revolves around the system time/date, hard and floppy disks, disk controller (IDE), keyboard, processor(s), memory, ISA/EISA/PCI expansion buses, system resources (IRQ, DMA, etc.), serial/parallel ports, and boot configuration. However, you should know some common features that you are more likely to find on a server than a workstation. After you access the BIOS settings in CMOS, use the menu system to navigate to various settings (see Table 3-7). The BIOS menu system varies from one manufacturer to the other, and a BIOS from a particular BIOS maker might be different from one machine to the next because server manufacturers often modify the menu system for uses specific to their computers.

Table 3-7 Common BIOS Settings

Feature	Description
Ultra DMA settings	Configures the high-throughput UDMA disk controller.
Processor	On a server, displays information about processor stepping (version) and L2 cache size. You can view the settings for each of the installed processors and test each one.
PCI bus mastering	Enables/disables devices as PCI bus masters and sets the number of clock cycles that a device can master on a PCI bus during a single transaction.
RAM testing	Selects the degree to which the system tests the installed memory (e.g., each 1 MB, 1 KB, or each byte boundary).
Memory scrubbing	Allows capable chipsets to automatically detect and correct single-bit memory errors.

Table 3-7 Common BIOS Settings (continued)

Feature	Description
Security	Allows the system to boot as usual, but prevents use of keyboard or mouse until the correct password is entered. Included on many servers, this secure boot mode also requires the password to boot from a floppy or CD-ROM. Some vendors secure the power and reset switches, which also require a password to use. You can usually configure a hot-key combination or countdown timer that places the server back in secure mode.
Management port	Specifies a serial port to which you can connect another device (a laptop, for example) to externally diagnose the server—even with the server turned off. You can configure management port settings such as whether a password is required to use the port, the connection type (port or modem), and so forth.
Logging	Logs system events in a small segment of nonvolatile memory.
Management interrupt	Allows system management to issue a **non-maskable interrupt** (**NMI**), which takes priority over standard interrupt requests. An NMI is useful to stop the system or issue a message in the event of critical events such as failing memory.
General hardware information	Displays hardware information such as part and serial numbers for the board, chassis, and system.
I2O drives	Allows you to specify the maximum number of I2O drives that will be assigned a DOS drive letter, usually one or four.

Most servers also include an additional management utility that provides similar management to that seen in the BIOS, except using a manufacturer-specific interface and settings. The settings are usually saved in BIOS, but settings can go beyond basic BIOS settings. For example, Intel's System Setup Utility also allows you to save **field replaceable unit (FRU)** information. (An FRU is a system with replaceable CPU, CMOS, CMOS battery, RAM, and RAM cache.) Typically, the management utility is a DOS-based utility run from floppy disk(s) or a CD-ROM. In some server types, however, these devices are not replaceable without sending the server to the manufacturer.

You should not use power management features in the NOS or in the BIOS. Servers generally need to be running 24/7, and you do not want to add latency to the server. However, power savings to the monitor are usually OK.

Chapter Summary

- A motherboard includes a front side bus, sometimes a back side bus, and three primary I/O expansion buses: Industry Standard Architecture (ISA), Extended ISA (EISA), and Peripheral Components Interconnect (PCI). Newer motherboards include only the PCI expansion bus. The speed of each bus is dependent upon the motherboard clock frequency.

- The front side bus connects to RAM, and the back side bus connects to L2 cache.
- Clock frequency (sometimes called the clock speed, cycle, or clock cycle) is the number of times in one second that an electrically charged quartz crystal located on the motherboard vibrates (oscillates). Clock frequency is measured in megahertz, a hertz equaling one cycle per second and "mega" meaning million.
- Using different buses in the system requires a way to divide the motherboard into separate parts. The chipset operates at the same speed as the motherboard clock and provides the boundary that both divides one bus from another (using a North Bridge and a South Bridge or the more current hub architecture) and controls the flow of bus data. Choosing a motherboard is mostly choosing a chipset.
- PC-based bus systems use what is known as a hierarchical bus, because several buses actually comprise the (collective) "bus," each running at different speeds and with the slower buses hierarchically structured beneath the faster buses. Dividing the bus into the front side bus, PCI bus, and ISA bus allows slower components to operate without negatively impacting the faster components.
- The accelerated hub architecture connects buses to the system bus independently through a dedicated hub interface to the PCI bus, yielding throughput of up to 266 MBps—twice as much throughput as 33 MHz PCI. The independent buses do not share the PCI bus—increasing PCI bus bandwidth available to PCI-connected devices.
- Devices connect to an ISA bus through an ISA expansion slot, which is 16 bits wide and accommodates both 16-bit devices and older 8-bit devices. The ISA bus only operates at 8.33 MHz and is capable of transfer speeds up to 8 MBps.
- The 32-bit EISA bus is backward compatible with older ISA devices and provides maximum bus bandwidth of about 33 MBps.
- The purpose of the PCI bus is to interface high-speed devices with the system bus so that slower devices do not bottleneck the system. The PCI bus also allows the processor to access L2 cache while simultaneously transferring data to and from other parts of the system.
- PCI-X utilizes 64 bits and up to 133 MHz, yielding a maximum bandwidth of 1064 MBps.
- In PCI architecture, hardware designers can create a device using bus mastering that is able to bypass the processor and directly access memory, resulting in an overall increase in processor performance. Bus mastering is actually a form of direct memory access (DMA) known as first-party DMA.
- The benefit of the PCI interrupt appears when no more ISA IRQ addresses are available. Then, with no more available IRQs, the PCI interrupt utilizes another PCI function known as PCI IRQ steering, in which the PCI interrupt assigns two or more PCI devices the same ISA IRQ.

- Although PCI hot swap (or hot plug) allows you to add, remove, or replace a device without first powering down the server, most servers still require you to power off the specific PCI slot using a button, switch, or software.
- The peer PCI bus is usually a server-specific function that both increases available PCI bandwidth and expands the number of PCI expansion cards from the usual limit of four with minimal impact on overall system bus bandwidth. This architecture usually involves dual peer PCI buses and two North Bridges, which connect to a primary PCI bus and a secondary PCI bus.
- I2O (Intelligent Input/Output) is an I/O design initiative that allows improved I/O performance via an I2O processor using the I2O driver model. With I2O, devices intelligently perform much of the processing function on their own instead of relying on the CPU.
- The design of the Accelerated Graphics Port (AGP) is to relieve the system bus and CPU of traffic and processing.
- Many servers come with AGP, particularly dual-processor machines that could just as easily serve as a high-end graphics workstation. However, high-end servers normally do not include AGP because it is not necessary.
- High-end servers usually include motherboard-integrated 1024 × 768 display adapters and only 256 colors.
- Processor speed is a measure in MHz of the number of opportunities per second that the processor can execute an instruction. The processor architecture design uses a multiplier methodology to provide the processor's speed.
- Cache memory is a small amount of memory that stores recently or frequently used program code or data, reducing latency. Cache memory appears in a number of places on the server including the hard disk, CD-ROM, and the processor. Processors use two types of cache memory: L1 (level 1) and L2 (level 2).
- L1 cache is a small amount of memory (usually 32–64 KB) and provides extremely fast access to its data because of its proximity to the processor and because it runs at the same speed as the processor itself—not the speed of the motherboard.
- L2 cache provides the same basic benefit as L1 cache, except that it is a larger cache, ranging from 256 KB to 2 MB. In the past, the L2 cache was not stored on the processor die, but was instead stored on a separate chip inside the processor housing. This orientation is known as discrete L2 cache. The data path used to access the L2 cache was called the back side bus, and it ran at half the processor speed.
- Most processors after the Pentium III 500 locate the L2 cache directly on the processor die (similar to the L1 cache) and run it at full processor speed. This is known as Advanced Transfer Cache (ATC), is 256 bits wide, and eliminates the need for a back side bus.

- Intel Xeon processors differ from standard Pentium II or III models in the following respects: type of enclosure, cache size, cache speed, addressable memory, and SMP.
- Intel Pentium II Xeon and later processors let the processor utilize 36 bits to address memory using Intel's Physical Address Extension (PAE) feature, allowing up to 64 GB addressable memory (2^36 = 68,719,476,736, or 64 GB).
- Xeon SMP configuration can use up to four processors, though by adding another processor bus often called a mezzanine bus, eight processors are possible, and many manufacturers engineer buses that can use up to 132 processors (though four or eight processors are more common).
- A dual-processor system is known as 2-way, four processors as 4-way, eight processors as 8-way, and so on.
- The 64-bit Intel Itanium using Intel's IA-64 technology represents a departure from the previous 32-bit x86 Intel architecture, and performs optimally with 64-bit operating systems (Windows 2000 will be ported to 64 bit) and applications.
- Because of the translation process from 32 to 64 bits, 32-bit applications will usually run more slowly on the IA-64 than on fast 32-bit Pentium III Xeon processors.
- With 64 address bits, the Itanium processor can address up to 18 billion GB.
- The IA-64 architecture uses Explicitly Parallel Instruction set Computing (EPIC), allowing the processor to simultaneously process as many as 20 operations.
- AMD decided to extend Intel's original x86 architecture in the Hammer design with AMD's new x86-64 architecture.
- 32-bit operating systems and applications can continue to run on the Hammer without complicated hardware emulation, resulting in minimal performance overhead.
- You will probably find SIMMs (single inline memory modules) on older servers, if at all. The physical SIMM module has gold or tin contacts at the bottom. Although the contact appears both on the front and back, it is really a single contact (hence the *single* in SIMM).
- The typical SIMM is 50 ns, 60 ns, or 70 ns, and older SIMMs can be 80 ns.
- Instead of providing only the exact location requested, EDO can send the entire row address so that subsequent references to the same row only require a column lookup, saving time.
- DIMMs (dual inline memory module) dramatically improve memory performance over the SIMM predecessor by expanding the module to 64 bits (nonparity) or 72 bits (parity or ECC). The contacts on both sides of the module are separate (hence the *dual* in DIMM).
- SDRAM operates at clock speed; if the system bus is 100 MHz, then SDRAM matches that frequency, which functionally operates at about 10 ns.

- RDRAM memory chips fit on a narrow, 16-bit-wide RIMM memory module.
- RDRAM provides extremely fast 800 MHz internal clock speed on a 400 MHz bus, because data is transferred twice during each clock cycle. This adds up to 1.6 GB throughput (16 bits × 800 MHz / 8 = 1.6 GB).
- Because of the unique data circuit of RDRAM, empty RIMM sockets must be filled with a C-RIMM, a device that has no memory but provides continuity to complete the memory data path.
- Double data rate SDRAM (DDR SDRAM) is the next generation of SDRAM, and also uses a 64-bit DIMM with future plans for a 128-bit DIMM.
- Interleaving allows memory access between two or more memory banks and/or boards to occur alternately, minimizing wait states. Memory must be installed in exactly the same configuration between banks/boards.
- The purpose of both buffered and registered memory is to re-drive (amplify) the signal entering the module.
- Buffered or registered modules also assist the chipset in handling the larger electrical load when the system has a lot of installed memory, allowing the module to include more memory chips, which is one reason that servers often use registered modules.
- Registered memory enacts a deliberate pause of one clock cycle in the module to ensure that all communication from the chipset arrives properly.
- Servers commonly use error correcting code (ECC) SDRAM. Although error correction is more expensive and involves a slight performance penalty, it is well worth it on a server, where data integrity is critical and other high-performing system components help make up for memory latency.
- If only a single bit error occurs, ECC can correct the error, but ECC cannot correct the more rare 2-, 3-, or 4-bit errors (multiple bit errors).
- The CMOS is a complimentary metal oxide semiconductor that includes a small amount of memory, the purpose of which is to store the BIOS settings.
- The power supply provides power to CMOS when the system is turned on, and a small, nonrechargeable, metal oxide battery (similar to a watch battery) supplies power when the server is turned off.
- The data stored in the CMOS is the BIOS settings. As its name implies, the BIOS is a series of input and output configuration settings for peripherals, adapters, and on-board components.
- Most of the time, when you turn on the power, the display tells you a specific key or keyboard combination to press in order to access the BIOS settings.
- When the system powers on, a procedure known as the POST (power on self-test) verifies functionality of the motherboard hardware.

- Protecting the BIOS for both the server and workstation is an important security precaution.
- The BIOS usually includes two levels of password protection: a password to access and change the BIOS configuration, intended to prevent the curious from so much as viewing the BIOS, and a password to boot the system. In the event that you cannot find a server's BIOS password, you have no choice but to reset the BIOS, which clears password settings in addition to any configuration settings.
- Reset the CMOS BIOS settings using jumpers or removing the battery.
- Most servers also include an additional management utility that provides similar management as seen in the BIOS, except using manufacturer-specific interface and settings.

KEY TERMS

Accelerated Graphics Port (AGP) — A high-speed graphics port that relieves the system bus and CPU of video-processing traffic.

accelerated hub architecture — Connects buses to the system bus independently through a dedicated hub interface to the PCI bus, yielding throughput of up to 266 MBps.

Advanced Transfer Cache (ATC) — L2 cache located on the processor die and running at full processor speed.

back side bus — The data path used to access L2 cache.

BIOS (basic input/output system) — A series of input and output configuration settings for peripherals, adapters, and on-board components.

buffered memory — Re-drives (amplifies) signals entering the memory module.

bus — Set of wires or printed circuits that provides the data path to and from the processor, memory, hard disk, adapters, and peripherals.

bus mastering — A technology that allows devices to bypass the processor and directly access memory, resulting in an overall increase in processor performance. Bus mastering devices can also communicate among themselves without processor intervention. Bus mastering is actually a form of direct memory access (DMA).

bus queue entries — A Pentium Xeon technology that holds outstanding bus and memory operations.

bus width — The number of individual data wires that transmit data. The more wires the component such as the motherboard has, the more data it can transmit in a given period of time.

cache memory — A small amount of memory that stores recently or frequently used program code or data.

chipset — Circuitry that provides motherboard features and organizes the various buses.

clock speed — The number of times in one second that the electrically charged quartz crystal located on the motherboard vibrates (oscillates). Also known as clock cycle, clock frequency, frequency, or cycle.

CMOS — Complimentary metal oxide semiconductor that includes a small amount of memory, the purpose of which is to store the BIOS settings.

compiler — Translates high-level programming language into the lowest language the computer can understand, machine language.

DDR SDRAM (double data rate SDRAM) — Transfers data twice per clock cycle, similar to RDRAM, but at a lower cost because DDR SDRAM is an open standard charging no royalties.

discrete L2 cache — L2 cache located inside the processor housing but not on the processor die.

dynamic RAM (DRAM) — Main memory referred to as dynamic because the information requires continuous electrical refresh, or else the data can become corrupt or lost.

error correcting code (ECC) — Circuitry on the memory chip that uses check bits to verify the integrity of memory and corrects single bit errors.

EEPROM (electrically erasable programmable read-only memory) — A chip that stores the BIOS programming. EEPROM has been mostly superceded by a similar memory known as flash BIOS.

Extended ISA (EISA) — An evolution of ISA, the EISA bus provides backward compatibility with older ISA devices and provides maximum bus bandwidth of about 33 MBps.

field replaceable unit (FRU) — A system with replaceable CPU, CMOS, CMOS battery, RAM, and RAM cache.

fill buffers — The interface between the CPU and main memory.

flash BIOS — BIOS memory that can be reprogrammed without having to remove the chip. Instead, you download and run a program that updates the BIOS.

front side bus — A 64-bit data pathway that the processor uses to communicate with L1 cache, main memory, and the graphics card through the North Bridge chipset.

Graphics Memory Controller Hub (GMCH) — Replaces the North Bridge in newer chipsets, providing higher data throughput.

hierarchical bus — Various portions of the bus running at different speeds, with the slower buses hierarchically structured beneath the faster buses.

Industry Standard Architecture (ISA) — A bus interface that connects ISA devices to the ISA bus, which is 16 bits wide and accommodates both 16-bit devices and older 8-bit devices. The ISA bus only operates at 8.33 MHz and is capable of transfer speeds up to 8 MBps.

Intelligent Input/Output (I2O) — An I/O design initiative that allows improved I/O performance via an I2O processor using the I2O driver model.

interleaving — A process that allows memory access between two or more memory banks and/or boards to occur alternately, minimizing wait states.

interrupt request (IRQ) — An electrical signal that obtains the CPU's attention in order to handle an event immediately, although the processor might queue the request behind other requests.

I/O Controller Hub (ICH) — Replaces the South Bridge in newer chipsets, allowing higher data throughput.

ISA bus — A 16-bit data pathway for slower expansion adapter cards and the floppy disk, mouse, keyboard, serial and parallel ports, and the BIOS via a Super I/O chip, which mitigates the need for a separate expansion card for each of the aforementioned items.

L1 cache — A small amount of memory (usually 32–64 KB) that provides extremely fast access to its data because of its proximity to the processor and because it runs at the same speed as the processor itself—not at the speed of the motherboard.

L2 cache — Provides the same basic benefits as L1 cache, but it is larger, ranging from 256 KB to 2 MB.

mezzanine bus — An add-on bus used to increase the number of processors in a single system.

motherboard — The heart of the computer, which attaches to the chassis and includes slots, sockets, and other connections for server components.

non-maskable interrupt (NMI) — An interrupt that takes priority over standard interrupt requests. An NMI is useful to stop the system or issue a message in event of critical events or failures such as failing memory.

North Bridge — A chipset element that divides the processor bus from the PCI bus and manages data traffic to and from the South Bridge, and components on the FSB and PCI bus.

null cable modem — A special cable that uses special crossed wires to simulate a modem presence, allowing data to travel between two hosts without an actual modem or network connection.

overclocking — Increasing the speed of the motherboard clock and/or the CPU to accelerate the clock speed, which can yield a performance increase. Not recommended on servers because of the higher risk associated with higher temperatures and a reduction in overall stability.

PCI hot swap (PCI hot plug) — The ability to add, remove, or replace PCI devices without first powering down the server.

PCI interrupts — Assignment of a designation to PCI devices that represent an actual ISA IRQ. The main benefit with PCI interrupts is that if no more IRQ addresses are available, PCI can use PCI steering to assign two or more PCI devices the same ISA IRQ.

PCI steering — Using PCI interrupts to assign two or more PCI devices the same ISA IRQ.

PCI-X (PCI-eXtended) — A 64-bit addendum to PCI 2.2 utilizing 64 bits and up to 133 MHz.

peer PCI bus — A bus architecture that increases available PCI bandwidth and expands the number of PCI expansion cards from the usual limit of four with a minimal impact on overall system bus bandwidth. This architecture usually involves dual peer PCI buses and two North Bridges, which connect to a primary PCI bus and a secondary PCI bus.

Peripheral Components Interface (PCI) bus — A 32-bit data pathway for high-speed I/O for expansion adapter cards, USB, and IDE ports. The CMOS and system clock also connect to the PCI bus. The PCI bus connects to both the North Bridge and the South Bridge.

Physical Address Extension (PAE) — Intel technology that allows the processor to utilize 36 bits to address up to 64 GB of memory.

POST (power-on self-test) — Verifies functionality of motherboard hardware.

RDRAM (Rambus DRAM) — Memory manufactured under license to Rambus. RDRAM is very fast, transferring data on both leading and trailing clock cycles.

registered memory — Memory that re-drives (amplifies) signals entering the memory module. Registered memory also enacts a deliberate pause of one clock cycle in the module to ensure that all communication from the chipset arrives properly. Registered memory is useful on heavily loaded server memory, and was designed for SIMMs containing 32 or more chips.

South Bridge — A chipset element that divides the PCI bus from the ISA bus.

synchronous dynamic RAM (SDRAM) — Memory that operates at system clock speed.

Review Questions

1. What is the main data transportation medium in the server?
   a. the PCI bus
   b. the motherboard
   c. the ISA bus
   d. ECC memory
2. How does the bus bit width affect data throughput?
   a. A wider bus results in less data throughput.
   b. Bus width does not affect data throughput.
   c. A wider bus results in more data throughput.
   d. A wider bus accepts higher voltage, allowing for overclocking.
3. If a server offers uptime of five nines, how many minutes per year might it be down?
   a. five minutes
   b. five increments of nine minutes each
   c. nine increments of five minutes each
   d. 59 minutes

4. How does clock frequency affect server performance?
   a. Each clock cycle represents a period of latency during which components cannot perform actions.
   b. Each clock cycle represents an opportunity for a component to perform an action.
   c. Only the processor, not clock frequency, affects performance.
   d. Slowing clock frequency means you should replace the CMOS battery soon.
5. What is the function of a chipset?
   a. manage data traffic and separate buses
   b. perform mathematical functions on behalf of the processor
   c. cache data to accelerate performance
   d. provide electrical continuity between the CPU and memory
6. What is a hierarchical bus?
   a. a bus that is the latest in a series of buses based on the same technology
   b. multiple motherboards in the same server
   c. a motherboard that uses a mezzanine bus to expand memory and/or processors
   d. the structuring of slower buses beneath faster buses
7. Which of the following motherboard buses is the fastest?
   a. ISA bus
   b. front side bus
   c. PCI bus
   d. USB
8. Which of the following buses are likely to be phased out on server boards?
   a. ISA bus
   b. PCI bus
   c. USB
   d. EISA bus
9. What is bus mastering?
   a. the ability of a device to bypass the processor and directly access memory
   b. a description of the system bus
   c. a description of the North Bridge's control over the system bus
   d. the ability of the GMCH to direct traffic between the processor and memory
10. How is a PCI interrupt beneficial?
    a. It allows PCI devices to have a priority when directing interrupt requests to the processor.
    b. It arbitrates between two PCI devices contending for processor time.

c. It allows PCI IRQ steering to assign two or more devices the same IRQ.

d. It divides a single PCI bus into two separate, faster PCI buses.

11. What do most servers require you to do before performing a PCI hot swap?
    a. Shut down the server power.
    b. Shut down power on the PCI slot.
    c. Reboot the server.
    d. Place the server in sleep mode.
12. What is the purpose of I2O? (Choose all that apply.)
    a. improve I/O performance
    b. simplify driver development
    c. provide a wider I/O bus
    d. provide a faster I/O bus
13. Why is AGP not included on some servers?
    a. AGP is incompatible with the Intel Xeon processor.
    b. Servers include a separate, independent bus to handle video.
    c. Servers do not require high-performance graphics functionality.
    d. AGP places too many demands on the processor.
14. What is the purpose of L1 and L2 cache?
    a. to supplement main memory in low memory conditions
    b. to store recently or frequently accessed data to reduce latency
    c. short-term storage in case of data corruption
    d. to cache data streams from the ISA bus
15. Which of the following characteristics does not differentiate the Intel Pentium III Xeon from the Pentium III?
    a. enclosure type
    b. cache size
    c. cache speed
    d. core clock speed
16. Which of the following is an advantage of the Intel Itanium?
    a. direct compatibility with 32-bit operating systems and applications
    b. Windows 2000 will be ported to 64 bits, taking advantage of the wider data path.
    c. support for up to eight processors
    d. cost-effective alternative to the Pentium III

17. What differentiates a SIMM from a DIMM?
    a. Contacts on either side of a SIMM are actually the same—on a DIMM, they are separate.
    b. You can only install one SIMM, but you install DIMMs in pairs.
    c. A DIMM has buffering, but a SIMM does not.
    d. A SIMM has registers, but a DIMM does not.
18. What makes SDRAM faster than standard DRAM?
    a. SDRAM operates at PCI speed.
    b. DRAM operates at bus speed.
    c. SDRAM operates at bus speed.
    d. They are the same speed.
19. Why is DDR SDRAM faster than standard SDRAM?
    a. DDR SDRAM transfers data twice per clock cycle.
    b. DDR SDRAM is not faster; it only has a wider data path.
    c. SDRAM can only run at half bus speed.
    d. SDRAM is only 32 bits wide.
20. Why do servers use ECC memory?
    a. increase performance
    b. increase latency
    c. increase data reliability
    d. purge memory in case of data errors
21. How does the BIOS relate to the CMOS?
    a. CMOS is the configuration programming that is stored in BIOS.
    b. CMOS stores BIOS configuration settings.
    c. BIOS and CMOS both provide I/O configuration settings for hardware.
    d. BIOS is an error-correcting utility for the CMOS.
22. What two benefits do BIOS passwords offer?
    a. ability to disable the power switch
    b. ability to prevent access to the operating system
    c. ability to prevent access to BIOS settings
    d. ability to prevent booting

HANDS-ON PROJECTS

Web links in projects were accurate at the time this book was published. If you notice discrepancies, look for similar links and follow the same general steps.

Project 3-1

In this project, you will observe a motherboard and answer several questions about it. Your instructor should have a motherboard for you to view. If possible, it should be at least a dual-processor motherboard.

Look at the motherboard provided by your instructor, and answer the following questions:

1. What chipset does the motherboard use?
2. How many processors can the motherboard accept?
3. What company makes the BIOS?
4. What kind of battery powers the CMOS?
5. Is there an AGP slot?

Project 3-2

In this project, you will access information on the Internet to describe information about Intel's server motherboards.

Using your Internet browser, access *http://support.intel.com/support/motherboards/server.* Intel makes bare motherboards for sale to system integrators (computer system manufacturers), listed under Boxed Server Boards. Intel also makes systems that are assembled motherboards, cases, and several components, listed under Server Platforms. Click **SBT2 Boxed Server Board** at the left. Answer the following questions:

1. How many processors does this board support?
2. What kinds of processors are supported?
3. What kind of memory is used?
4. What is the maximum amount of memory supported?
5. Click the **Supported Processors** chart. What is the fastest supported processor?
6. Click the **Technical Notes** link, and open the document titled **Microsoft Windows 2000 Advanced Server and 4GB or greater memory installed**. What must you do in order for Windows 2000 Advanced Server to recognize 4 GB or more of RAM?

Project 3-3

In this project, you find out how to access your BIOS settings and view some of its configurations.

1. Boot your server and observe what appears on the screen as the POST takes place. List what you see as best you can. You might need to reboot a couple of times (or try pressing the Pause key) as some of the information flashes quickly on screen.
2. What key would you press to access BIOS settings?

Project 3-4

In this project, you use the BIOS menu to find configuration information.

1. Boot your server and access the BIOS configurations.
2. Using the menu system, answer the following questions:
   a. How much memory is installed in the computer?
   b. How large is the hard disk?
   c. In what order does the BIOS search for bootable media?
3. Attempt to exit the BIOS configuration. You are prompted to save your settings. Choose to discard the changes.

Project 3-5

In this project, you will see a visual demonstration of how the Intel chipset architecture works.

1. Access the Intel web site at *http://developer.intel.com/design/chipsets.*
2. Under **Server**, click the link to the **Intel 840 Chipset**.
3. Under the **Spotlight** heading, click **840 Chipset Architecture Demo**, and play the video. Be sure to click each of the links on the right-hand side of the window.
4. Is the chipset a North Bridge/South Bridge? If not, what chipset is it?
5. How many processors can the chipset use?
6. What kind of processors does it use?
7. What type of memory does this chipset use?

Project 3-6

In this project, you learn how to quickly access additional information about memory.

1. Access the Kingston Technology Company web site at *www.kingston.com.*
2. Click the link for **Educational Tools**.
3. Click the link for **Memory Bits**.

4. Click several topics at the left to research more about memory.
5. Return to the home page, and click the **Memory Configurator** link. Locate the correct memory for the Compaq ProLiant 8500. What is the largest single memory module Kingston makes for this server? How much does it cost? (Click **Add to Basket** to find out.)

CASE PROJECTS

1. KidHelp, a nonprofit charity, asks you to donate your expertise to help them with their growing organization's network. (You are a notable humanitarian, so of course you agree to help.) The network has grown from a peer-to-peer network of six users to about 50 regular users. Somewhere in the growth path, somebody donated a NetWare 3.11 server, the main purpose of which is as a file and print server, which is configured as follows:

   - One Classic Pentium 60 MHz
   - 16 MB RAM
   - One 780 MB hard disk
   - One CD-ROM drive

   KidHelp has biannual fundraisers in which 250 volunteers answer phones and write down donation information, which someone later enters into a spreadsheet. A larger charity has donated their 500 MB SQL database of donors to KidHelp, and KidHelp wants to add their own donors to the SQL database and use it during the next fundraiser. KidHelp has asked you to upgrade the hardware and NOS with the following objectives:

   - Upgrade somewhat ahead of current needs to increase the length of time necessary for the next upgrade.
   - Use two processors.
   - Utilize a motherboard with a fast, high-throughput network and SCSI adapters.
   - Utilize a memory solution that would allow for more memory chips per module so that as the database grows, it can still be loaded completely into memory. Also, the memory should automatically correct errors.
   - Optionally, upgrade the NOS to another version of NetWare or to Windows 2000.

   What kind of server hardware would you suggest?

2. Your pager alerts you that a network card has failed on one of your file and print servers. You want to replace the NIC immediately, except that because you recently used most of your spare NICs in some new servers, you only have one left. Unfortunately, this NIC is from a different manufacturer than the original one, and usually requests a different IRQ—one that is in use by the AGP video card. Also, you would like to replace the NIC without shutting off the server. What steps should you take to replace the NIC? What PCI technologies help to make this NIC replacement smooth and why?